PCT10

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/089,057

DATE: 02/14/2003
TIME: 12:41:31

Input Set : A:\EP.txt
Output Set: N:\CRF4\02142003\J089057.raw

```
 3 <110> APPLICANT: HIRANO, SEIKO
 4        NONAKA, GEN
 5        MATSUZAKI, YUMI
 6.       AKIYOSHI, NAOKI
 7        NAKAMURA, KANAE
 8        KIMURA, EIICHIRO
 9        OSUMI, TSUYOSHI
10        MATSUI, KAZUHIKO ·
11        KAWAHARA, YOSHIO
12        KURAHASHI, OSAMU
13        NAKAMATSU, TSUYOSHI
14        SUGIMOTO, SHINICHI
16 <120> TITLE OF INVENTION: GENES FOR HEAT RESISTANT ENZYMES OF AMINO ACID BIOSYNTHETIC
PATHWAY
17        DERIVED FROM THERMOPHILIC CORYNEFORM BACTERIA
19 <130> FILE REFERENCE: 221519US0PCT
21 <140> CURRENT APPLICATION NUMBER: 10/089,057
22 <141> CURRENT FILING DATE: 2002-04-03
24 <150> PRIOR APPLICATION NUMBER: PCT/JP00/6913
25 <151> PRIOR FILING DATE: 2000-10-04
27 <150> PRIOR APPLICATION NUMBER: JP11-282716
28 <151> PRIOR FILING DATE: 1999-10-04
30 <150> PRIOR APPLICATION NUMBER: JP11-311147
31 <151> PRIOR FILING DATE: 1999-11-01
33 <150> PRIOR APPLICATION NUMBER: JP 2000-120687
34 <151> PRIOR FILING DATE: 2000-04-21
36 <160> NUMBER OF SEQ ID NOS: 129
38 <170> SOFTWARE: PatentIn version 3.1
40 <210> SEQ ID NO: 1
41 <211> LENGTH: 1980
42 <212> TYPE: DNA
43 <213> ORGANISM: Corynebacterium thermoaminogenes
45 <220> FEATURE:
46 <221> NAME/KEY: CDS
47 <222> LOCATION: (577)..(1869)
48 <223> OTHER INFORMATION:
W--> 50 <400> 1
51 tgcattccac cgacggtcac gcgttcggtc ttgtcagcgg cgtcaatctg ctgatggttc    60
53 atgcaaagct ccttcgaagc aagagatcgg gtgtgtgcgg gcacctatcg ggggaagccc   120
55 tcgctgcgcc ccagggggag ctggcgatgt gaccaggtta agtgataacc atcaccttgc   180
57 caatgggttt gcgaacttta ccgtgacgct accccgctt ttgtttgatc acaccagctc    240
59 gaaggctgtc gcttttccga agatgcacgt gaagtggcaa atccttccca cccgaggttt    300
61 tcccagtaca aacgtactag tgatgaggat cacggggaac attgtggaga ttgcactttg    360
63 caatatttgc aaagggggtg actacccccg cgcaaaactt aaaaacccaa atccgttgac    420
```

RAW SEQUENCE LISTING   DATE: 02/14/2003
PATENT APPLICATION: **US/10/089,057**   TIME: 12:41:31

Input Set : **A:\EP.txt**
Output Set: **N:\CRF4\02142003\J089057.raw**

```
65  ggacccatgc ccgatgaagc aatgtgtgaa gcacgccacc ggaacacagg ttgtggatca    480
-67 ctcaccatga tgtgggggat tcgcatcaca cagtgtgcag ggcggcacct ctaccgaatg    540
69  cgccttacag cagcaccaag aagaagtgac tcttag atg tca aac gtt gga acg      594
70                                          Met Ser Asn Val Gly Thr
71                                          1                   5
73  cca cgt acc gca cag gaa atc cag cag gat tgg gac acc aac cca cgc      642
74  Pro Arg Thr Ala Gln Glu Ile Gln Gln Asp Trp Asp Thr Asn Pro Arg
75              10                  15                  20
77  tgg aac gga atc acc cgc gac tac acc gct gag cag gta gct gag ctc      690
78  Trp Asn Gly Ile Thr Arg Asp Tyr Thr Ala Glu Gln Val Ala Glu Leu
79          25                  30                  35
81  cag ggc agc gtc gtc gag gag cac acc ctc gca aag cgc ggc gcc gag      738
82  Gln Gly Ser Val Val Glu Glu His Thr Leu Ala Lys Arg Gly Ala Glu
83      40                  45                  50
85  atc ctg tgg gat gca gtt tcc gca gag ggc gac gac tac atc aac gca      786
86  Ile Leu Trp Asp Ala Val Ser Ala Glu Gly Asp Asp Tyr Ile Asn Ala
87  55                  60                  65                  70
89  ctg ggc gcc ctt acc ggt aac cag gct gtc cag cag gtc cgt gcc ggc      834
90  Leu Gly Ala Leu Thr Gly Asn Gln Ala Val Gln Gln Val Arg Ala Gly
91                  75                  80                  85
93  ctg aag gct gtc tac ctc tcc ggc tgg cag gtc gca ggt gac gcc aac      882
94  Leu Lys Ala Val Tyr Leu Ser Gly Trp Gln Val Ala Gly Asp Ala Asn
95              90                  95                  100
97  ctc gcc ggt cac acc tac ccc gac cag tcc ctg tac ccg gcg aac tcc      930
98  Leu Ala Gly His Thr Tyr Pro Asp Gln Ser Leu Tyr Pro Ala Asn Ser
99              105                 110                 115
101 gtc ccg aac gtt gtc cgt cgc atc aac aac gca ctg ctg cgc gcc gat      978
102 Val Pro Asn Val Val Arg Arg Ile Asn Asn Ala Leu Leu Arg Ala Asp
103         120                 125                 130
105 gag atc gca cgc gtc gag ggt gac acc tcc gtc gac aac tgg ctc gtc     1026
106 Glu Ile Ala Arg Val Glu Gly Asp Thr Ser Val Asp Asn Trp Leu Val
107 135                 140                 145                 150
109 ccg atc gtc gcc gac ggc gag gcc ggc ttc ggt ggc gcc ctc aac gtc     1074
110 Pro Ile Val Ala Asp Gly Glu Ala Gly Phe Gly Gly Ala Leu Asn Val
111                 155                 160                 165
113 tac gag ctc cag aag ggc atg atc acc gct ggt gcc gca ggc acc cac     1122
114 Tyr Glu Leu Gln Lys Gly Met Ile Thr Ala Gly Ala Ala Gly Thr His
115                 170                 175                 180
117 tgg gag gat cag ctc gct tcc gag aag aag tgt ggc cac ctc ggt ggc     1170
118 Trp Glu Asp Gln Leu Ala Ser Glu Lys Lys Cys Gly His Leu Gly Gly
119                 185                 190                 195
121 aag gtc ctc atc ccg acc cag cag cac atc cgc acc ctg aac tcc gcc     1218
122 Lys Val Leu Ile Pro Thr Gln Gln His Ile Arg Thr Leu Asn Ser Ala
123                 200                 205                 210
125 cgc ctg gca gct gac gtg gcc aac acc ccg acc gtc gtc atc gcc cgc     1266
126 Arg Leu Ala Ala Asp Val Ala Asn Thr Pro Thr Val Val Ile Ala Arg
127 215                 220                 225                 230
129 acc gac gca gag gcc gcc acc ctg atc acc tct gat gtt gat gag cgc     1314
130 Thr Asp Ala Glu Ala Ala Thr Leu Ile Thr Ser Asp Val Asp Glu Arg
```

RAW SEQUENCE LISTING   DATE: 02/14/2003
PATENT APPLICATION: US/10/089,057   TIME: 12:41:31

Input Set : A:\EP.txt
Output Set: N:\CRF4\02142003\J089057.raw

```
131              235              240              245
-133 gac cgc cca ttc atc acc ggc gag cgc acc gcc gag ggc tac tac cac   1362
134 Asp Arg Pro Phe Ile Thr Gly Glu Arg Thr Ala Glu Gly Tyr Tyr His
135              250              255              260
137 gtc aag ccg ggt ctc gag ccc tgc atc gca cgt gcg aag tcc tac gct   1410
138 Val Lys Pro Gly Leu Glu Pro Cys Ile Ala Arg Ala Lys Ser Tyr Ala
139              265              270              275
141 ccc tac gca gac atg atc tgg atg gag acc ggc acc cct gac ctc gag   1458
142 Pro Tyr Ala Asp Met Ile Trp Met Glu Thr Gly Thr Pro Asp Leu Glu
143.             280              285              290
145 ctg gcc aag aag ttc gcc gag ggc gtc cgc agc gag ttc ccg gac cag   1506
146 Leu Ala Lys Lys Phe Ala Glu Gly Val Arg Ser Glu Phe Pro Asp Gln
147 295              300              305              310
149 ctg ctg tcc tac aac tgc tcc ccg tcc ttc aac tgg tct gca cac ctc   1554
150 Leu Leu Ser Tyr Asn Cys Ser Pro Ser Phe Asn Trp Ser Ala His Leu
151              315              320              325
153 gag gcc gac gag atc gct aag ttc cag aag gaa ctg ggt gcc atg ggc   1602
154 Glu Ala Asp Glu Ile Ala Lys Phe Gln Lys Glu Leu Gly Ala Met Gly
155              330              335              340
157 ttc aag ttc cag ttc atc acc ctg gct ggc ttc cac tcc ctc aac tac   1650
158 Phe Lys Phe Gln Phe Ile Thr Leu Ala Gly Phe His Ser Leu Asn Tyr
159              345              350              355
161 ggt atg ttc gac ctg gct tac ggc tac gcc cgt gaa ggc atg ccc gcc   1698
162 Gly Met Phe Asp Leu Ala Tyr Gly Tyr Ala Arg Glu Gly Met Pro Ala
163      360              365              370
165 ttc gtc gac ctg cag aac cgt gag ttc aag gca gct gag gag cgc ggc   1746
166 Phe Val Asp Leu Gln Asn Arg Glu Phe Lys Ala Ala Glu Glu Arg Gly
167 375              380              385              390
169 ttc acc gcc gtc aag cac cag cgt gag gtc ggc gcc ggc tac ttc gac   1794
170 Phe Thr Ala Val Lys His Gln Arg Glu Val Gly Ala Gly Tyr Phe Asp
171              395              400              405
173 acc atc gcc acc acc gtt gac ccg aac tcc tcc acc acc gcg ctg aag   1842
174 Thr Ile Ala Thr Thr Val Asp Pro Asn Ser Ser Thr Thr Ala Leu Lys
175              410              415              420
177 ggt tcc acc gag gaa tgc cag ttc cac taggaaccac ctgatgcggt   1889
178 Gly Ser Thr Glu Glu Cys Gln Phe His
179      425              430
181 gccgtatggc ctgacggcac cgcccctccc tttgcactcc agtactcctt tgtgcacatc   1949
183 ggccatctcc acccgcgcg ccccgccacc t   1980
186 <210> SEQ ID NO: 2
187 <211> LENGTH: 431
188 <212> TYPE: PRT
189 <213> ORGANISM: Corynebacterium thermoaminogenes
191 <400> SEQUENCE: 2
193 Met Ser Asn Val Gly Thr Pro Arg Thr Ala Gln Glu Ile Gln Gln Asp
194 1             5              10             15
197 Trp Asp Thr Asn Pro Arg Trp Asn Gly Ile Thr Arg Asp Tyr Thr Ala
198      20             25             30
201 Glu Gln Val Ala Glu Leu Gln Gly Ser Val Val Glu Glu His Thr Leu
```

RAW SEQUENCE LISTING   DATE: 02/14/2003
PATENT APPLICATION: US/10/089,057   TIME: 12:41:31

Input Set : A:\EP.txt
Output Set: N:\CRF4\02142003\J089057.raw

```
202        35                40                45
-205 Ala Lys Arg Gly Ala Glu Ile Leu Trp Asp Ala Val Ser Ala Glu Gly
206     50                55                60
209 Asp Asp Tyr Ile Asn Ala Leu Gly Ala Leu Thr Gly Asn Gln Ala Val
210 65                70                75                80
213 Gln Gln Val Arg Ala Gly Leu Lys Ala Val Tyr Leu Ser Gly Trp Gln
214            85                90                95
217 Val Ala Gly Asp Ala Asn Leu Ala Gly His Thr Tyr Pro Asp Gln Ser
218           100               105               110
221.Leu Tyr Pro Ala Asn Ser Val Pro Asn Val Val Arg Arg Ile Asn Asn
222          115               120               125
225 Ala Leu Leu Arg Ala Asp Glu Ile Ala Arg Val Glu Gly Asp Thr Ser
226      130               135               140
229 Val Asp Asn Trp Leu Val Pro Ile Val Ala Asp Gly Glu Ala Gly Phe
230 145               150               155               160
233 Gly Gly Ala Leu Asn Val Tyr Glu Leu Gln Lys Gly Met Ile Thr Ala
234           165               170               175
237 Gly Ala Ala Gly Thr His Trp Glu Asp Gln Leu Ala Ser Glu Lys Lys
238           180               185               190
241 Cys Gly His Leu Gly Gly Lys Val Leu Ile Pro Thr Gln Gln His Ile
242      195               200               205
245 Arg Thr Leu Asn Ser Ala Arg Leu Ala Ala Asp Val Ala Asn Thr Pro
246      210               215               220
249 Thr Val Val Ile Ala Arg Thr Asp Ala Glu Ala Ala Thr Leu Ile Thr
250 225               230               235               240
253 Ser Asp Val Asp Glu Arg Asp Arg Pro Phe Ile Thr Gly Glu Arg Thr
254           245               250               255
257 Ala Glu Gly Tyr Tyr His Val Lys Pro Gly Leu Glu Pro Cys Ile Ala
258           260               265               270
261 Arg Ala Lys Ser Tyr Ala Pro Tyr Ala Asp Met Ile Trp Met Glu Thr
262      275               280               285
265 Gly Thr Pro Asp Leu Glu Leu Ala Lys Lys Phe Ala Glu Gly Val Arg
266      290               295               300
269 Ser Glu Phe Pro Asp Gln Leu Leu Ser Tyr Asn Cys Ser Pro Ser Phe
270 305               310               315               320
273 Asn Trp Ser Ala His Leu Glu Ala Asp Glu Ile Ala Lys Phe Gln Lys
274           325               330               335
277 Glu Leu Gly Ala Met Gly Phe Lys Phe Gln Phe Ile Thr Leu Ala Gly
278           340               345               350
281 Phe His Ser Leu Asn Tyr Gly Met Phe Asp Leu Ala Tyr Gly Tyr Ala
282      355               360               365
285 Arg Glu Gly Met Pro Ala Phe Val Asp Leu Gln Asn Arg Glu Phe Lys
286      370               375               380
289 Ala Ala Glu Glu Arg Gly Phe Thr Ala Val Lys His Gln Arg Glu Val
290 385               390               395               400
293 Gly Ala Gly Tyr Phe Asp Thr Ile Ala Thr Thr Val Asp Pro Asn Ser
294           405               410               415
297 Ser Thr Thr Ala Leu Lys Gly Ser Thr Glu Glu Cys Gln Phe His
298           420               425               430
```

RAW SEQUENCE LISTING
PATENT APPLICATION: **US/10/089,057**

DATE: 02/14/2003
TIME: 12:41:31

Input Set : **A:\EP.txt**
Output Set: **N:\CRF4\02142003\J089057.raw**

```
301 <210> SEQ ID NO: 3
302 <211> LENGTH: 2381
303 <212> TYPE: DNA
304 <213> ORGANISM: Corynebacterium thermoaminogenes
306 <220> FEATURE:
307 <221> NAME/KEY: CDS
308 <222> LOCATION: (577)..(2349)
309 <223> OTHER INFORMATION:
```

W--> 311 <400> 3

```
312 agcaggccgt gttgccgaac ggcaacttcc gcagccgcaa ggagatcgag gaggtgtact      60
314 cgcacctcaa ccctgccgag gacaccgtgg tgtactgccg cgtgggtgac cgcgcggccc     120
316 acacctggtt cgtgttgaag tacctgctgg ggtttgaaaa cgtccgcaac tatgacggtt     180
318 cctggtccga gtggggcaac atggtgcgca tgcccatcgt ccagggtgat gagccgggct     240
320 cactctagtc accccggggt cacctccctg gtcacccccg taccctcccg ggtacacccc     300
322 ggggacgggg tgtgacctgg atctcccctg catgtggaca ccgggaaact ttgcctggga     360
324 aatgaccatc cagtaccgta atgcgggtat gttaacgcgg tcacagggta caccagaatc     420
326 cggatcgtct aacccccctta gcgggattcg ctaaaagatc accgagttag tgtgcaagaa     480
328 taatgctgat cgcaggggca ctgtcatacg ctgtcatgca gtcaatgaac agtgcggtgc     540
330 tctgtcgtga agaaaatcaa aaccaggagg gtttta gtg tca gtc gag acc agg      594
332                                           Val Ser Val Glu Thr Arg
                                              1               5
334 aag atc acc aag gta ctt gtc gcc aac cgt ggt gaa atc gca atc cgt      642
335 Lys Ile Thr Lys Val Leu Val Ala Asn Arg Gly Glu Ile Ala Ile Arg
336             10              15                  20
338 gtt ttc cgc gca gca cgg gat gaa ggc atc gcc tct gtc gcc gtc tac      690
339 Val Phe Arg Ala Ala Arg Asp Glu Gly Ile Ala Ser Val Ala Val Tyr
340         25              30                  35
342 gcg gag ccg gac gca gat gcc cct ttc gtc gag tat gcc gat gag gcc      738
343 Ala Glu Pro Asp Ala Asp Ala Pro Phe Val Glu Tyr Ala Asp Glu Ala
344     40              45                  50
346 ttc gca ctc ggt ggc cag act tcc gca gag tcc tac ctc gtc att gac      786
347 Phe Ala Leu Gly Gly Gln Thr Ser Ala Glu Ser Tyr Leu Val Ile Asp
348 55              60                  65                  70
350 aag atc att gac gca gca cgc aag tcc ggt gca gac gct gtc cac ccc      834
351 Lys Ile Ile Asp Ala Ala Arg Lys Ser Gly Ala Asp Ala Val His Pro
352             75                  80                  85
354 ggc tac ggc ttc ctc gcc gag aac gcc gat ttc gct gaa gct gtc atc      882
355 Gly Tyr Gly Phe Leu Ala Glu Asn Ala Asp Phe Ala Glu Ala Val Ile
356             90                  95                  100
358 aac gag ggc ctg atc tgg atc gga cca tcc cct gag tcc atc cgt tcc      930
359 Asn Glu Gly Leu Ile Trp Ile Gly Pro Ser Pro Glu Ser Ile Arg Ser
360             105                 110                 115
362 ctc ggt gac aag gtc acc gca cgc cac atc gcc aac aac gcc aac gca      978
363 Leu Gly Asp Lys Val Thr Ala Arg His Ile Ala Asn Asn Ala Asn Ala
364     120                 125                 130
366 ccg atg gca ccg ggc acc aag gag cct gtc aag gac gcc gct gag gtt     1026
367 Pro Met Ala Pro Gly Thr Lys Glu Pro Val Lys Asp Ala Ala Glu Val
368 135                 140                 145                 150
370 gtc gcc ttc gcc gag gag ttc ggt ctc ccc atc gcc atc aag gct gcc     1074
```

RAW SEQUENCE LISTING ERROR SUMMARY   DATE: 02/14/2003
PATENT APPLICATION: **US/10/089,057**   TIME: 12:41:32

Input Set : **A:\EP.txt**
Output Set: **N:\CRF4\02142003\J089057.raw**

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:66; N Pos. 3
Seq#:67; N Pos. 18
Seq#:76; N Pos. 3,6,9
Seq#:77; N Pos. 3,9,18
Seq#:104; N Pos. 9

VERIFICATION SUMMARY   DATE: 02/14/2003
PATENT APPLICATION: **US/10/089,057**   TIME: 12:41:32

Input Set : **A:\EP.txt**
Output Set: **N:\CRF4\02142003\J089057.raw**

```
L:50 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:1,Line#:48
L:311 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:3,Line#:309
L:650 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:5,Line#:648
L:961 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:7,Line#:959
L:1270 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:9,Line#:1268
L:1487 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:11,Line#:1485
L:1594 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:13,Line#:1592
L:1723 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:16,Line#:1716
L:1728 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:16,Line#:1721
L:1733 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:16,Line#:1726
L:1733 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:16,Line#:1731
L:2353 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:21,Line#:2351
L:2866 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:23,Line#:2864
L:3664 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:46,Line#:3662
L:4165 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:48,Line#:4163
L:4680 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:50,Line#:4678
L:5095 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:52,Line#:5093
L:5384 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:54,Line#:5382
L:6172 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:66 after pos.:0
L:6190 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:67 after pos.:0
L:6316 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:76 after pos.:0
L:6346 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:77 after pos.:0
L:6623 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:100,Line#:6621
L:6866 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:102,Line#:6864
L:7114 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:104 after pos.:0
L:7187 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:110,Line#:7185
L:7422 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:112,Line#:7420
L:7665 M:258 W: Mandatory Feature missing, <223> Blank for SEQ#:114,Line#:7663
```